On the Convergence of Asynchronous Parallel Pattern Search

TAMARA G. KOLDA
Sandia National Labs

VIRGINIA TORCZON
College of William & Mary

Asynchronous Parallel Pattern Search

APPS (Hough, K., Torczon; 2000) is pattern search method for solving the unconstrained optimization problem

$$\min f(x), \quad x \in \Re^n.$$

Development of APPS was motivated by the need to remove the synchronization point in each iteration of standard parallel pattern search.

The theoretical presentation of the method is quite different than the practical presentation.

Theoretical Framework

Processors:	$\mathcal{P} = \{1, \ldots, p\}$
Search Directions:	$\mathcal{D} = \{d_1, \ldots, d_p\}$
Global Time Index:	$\mathcal{T} = \{0, 1, \ldots\}$
Best Known Point:	x_i^t (at time t for process i)
Step Length:	Δ_i^t (at time t for process i)
Change Indices:	$\mathcal{T}_i$ (for process i)

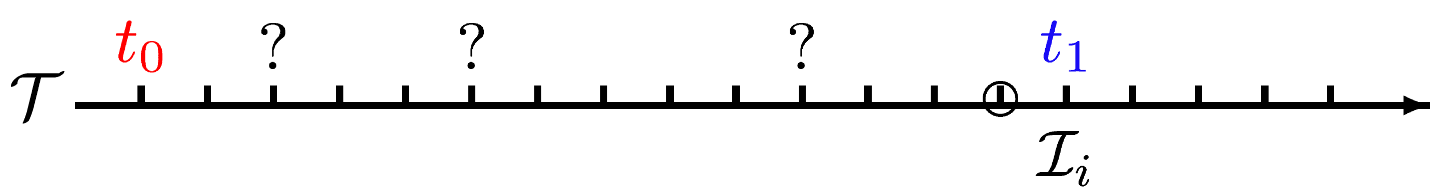
At time 0, each process starts a function evaluation at the point $x_i^0 + \Delta_i^0 \, d_i$. It is assumed that $f(x_i^0)$ is known.

Internal Successful Iterates

$\mathcal{I}_i$ = Set of internal successful iterates for process i

t_0 = Time function evaluation begins on process i

t_1 = Time function evaluation completes on process i

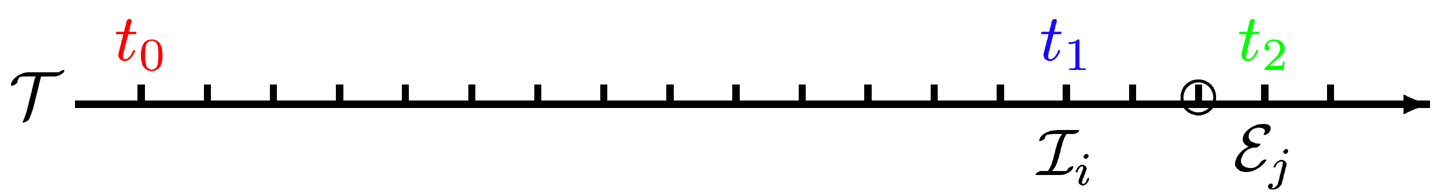


$$
\begin{array}{ll}
\text{If} & f\left(x_i^{t_0} + \Delta_i^{t_0}\, d_i\right) < f\left(x_i^{t_1 - 1}\right), \\
\text{then} & x_i^{t_1} = x_i^{t_0} + \Delta_i^{t_0}\, d_i, \\
\text{and} & \Delta_i^{t_1} = \lambda_i^{t_1} \Delta_i^{t_0} \ (\Delta_{\min} \le \Delta_i^t \le \Delta_{\max}).
\end{array}
$$

External Successful Iterates

$\mathcal{E}_i$ = Set of external successful iterates for process i

t_2 = Time $x_i^{t_1}$ is communicated to process j



If

$$\begin{aligned} f\left(x_i^{t_1}\right) &< f\left(x_j^{t_2-1}\right), \quad \text{or} \\ f\left(x_i^{t_1}\right) &= f\left(x_j^{t_2-1}\right) \quad \text{and} \quad x_i^{t_1} \prec x_j^{t_2-1}, \end{aligned}$$

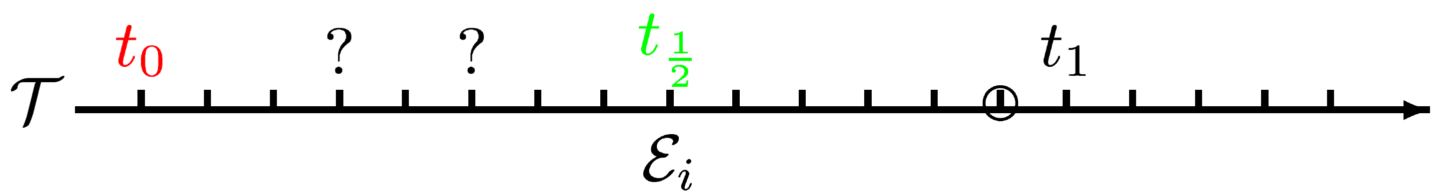
then

$$x_j^{t_2} = x_i^{t_1} \quad \text{and} \quad \Delta_j^{t_2} = \Delta_i^{t_1},$$

External Successful Iterate

$t_{\frac{1}{2}}$ = Time of external success on process i

t_1 = Time function evaluation completes on process i



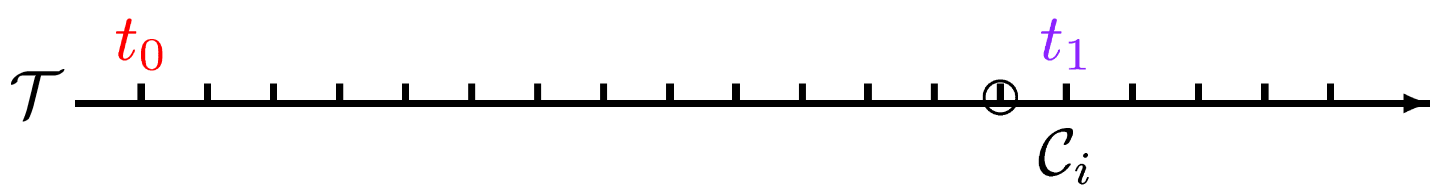
If

$$x_i^{t_{\frac{1}{2}}} = x_i^{t_1 - 1} \quad \text{and} \quad f\left(x_i^{t_1-1}\right) \leq f\left(x_i^{t_0} + \Delta_i^{t_0}\, d_i\right)$$

then start new function evaluations with the information from the external success.

Contraction Iterates

$\mathcal{C}_i$ = Set of contraction iterates for process i



If

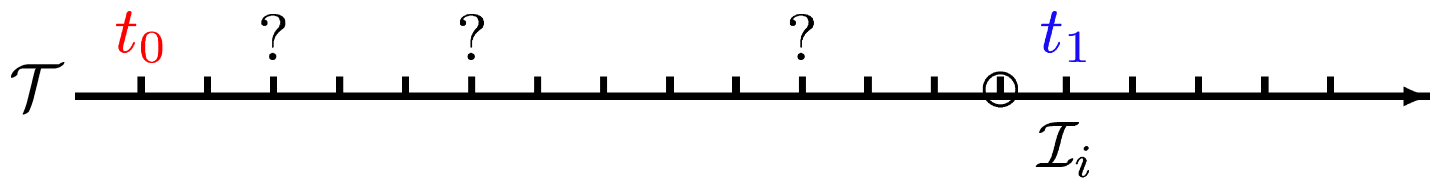
$$x_i^{t_0} = x_i^{t_1 - 1} \quad \text{and} \quad f\left(x_i^{t_1-1}\right) \leq f\left(x_i^{t_0} + \Delta_i^{t_0}\, d_i\right)$$

then

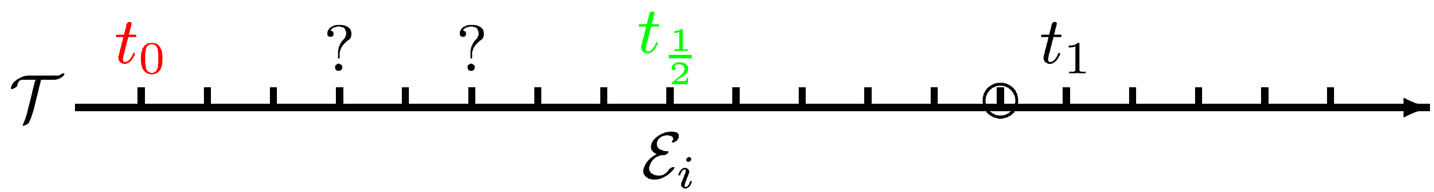
$$\Delta_i^{t_1} = \theta_i^{t_1} \Delta_i^{t_0} \quad \text{where} \quad \theta_i^t \in [\theta_{\min}, \theta_{\max}] \subset (0, 1)$$

End of Function Evaluation

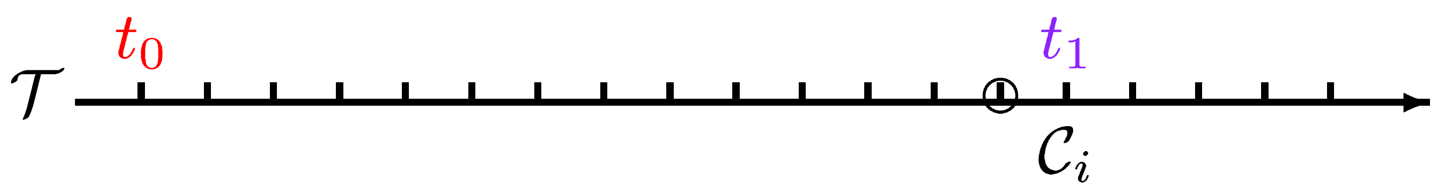
Case I: Internal Success



Case II: (Interim) External Success



Case III: Contraction



Update Formulas

$\omega_i(t)$ = The generating process index

$\tau_i(t)$ = Start time index

$\nu_i(t)$ = Completion time index

$$x_i^t = \begin{cases} x_{\omega_i(t)}^{\tau_i(t)} + \Delta_{\omega_i(t)}^{\tau_i(t)}\, d_{\omega_i(t)}, & \text{if } t \in \mathcal{S}_i (\equiv \mathcal{I}_i \cup \mathcal{E}_i), \\ x_i^{t-1}, & \text{otherwise.} \end{cases}$$

$$\Delta_i^t = \begin{cases} \lambda_{\omega_i(t)}^{\nu_i(t)}\, \Delta_{\omega_i(t)}^{\tau_i(t)}, & \text{if } t \in \mathcal{S}_i, \\ \theta_i^t \Delta_i^{t-1}, & \text{if } t \in \mathcal{C}_i, \\ \Delta_i^{t-1}, & \text{otherwise.} \end{cases}$$

Lemma: $\mathcal{T}_i \ (= \mathcal{S}_i \cup \mathcal{C}_i)$ is infinite.

Reducing the Step Length

Goal: $\liminf_{t \to +\infty} \Delta_j^t = 0$ for all $j \in \mathcal{P}$.

Lemma: If $\mathcal{S}_i$ is finite for some $i \in \mathcal{P}$, then

$$\lim_{t \to +\infty} \Delta_i^t = 0.$$

Lemma: If $\mathcal{S}_i$ is finite for some i, then $\mathcal{S}_j$ is finite for all $j \in \mathcal{P}$.

$\Rightarrow$ *Finite case is trivial.*

Corollary: If $\mathcal{S}_i$ is infinite for some $i \in \mathcal{P}$, then $\mathcal{S}_j$ is infinite for all $j \in \mathcal{P}$.

One Step Length Goes To Zero

Lemma: Assume the level set $\mathcal{L}(x_0)$ is compact. Suppose $\mathcal{S}_j$ is infinite for all $j \in \mathcal{P}$, then there exists $i \in \mathcal{P}$ such that

$$\liminf_{\substack{t \to +\infty \\ t \in \mathcal{S}_i}} \Delta_{\omega_i(t)}^{\tau_i(t)} = 0.$$

Finite rational lattice argument — Torczon, 1997.

Corollary: Suppose $\mathcal{S}_j$ is infinite for all $j \in \mathcal{P}$, then there exists $i \in \mathcal{P}$ such that

$$\liminf_{t \to +\infty} \Delta_i^t = 0.$$

All Step Lengths Go To Zero

Let process i be such that

$$\liminf_{t \to +\infty} \Delta_i^t = 0,$$

Key: *The supremum of the time between successful iterates on process i goes to $+\infty$.*

So, on any other process j, the supremum of the number of contractions between successful iterates is also going to $+\infty$.

Theorem 1: $\liminf_{t \to +\infty} \Delta_j^t = 0$ for all $j \in \mathcal{P}$.

Accumulation Point

Goal: There exists $\hat{x} \ni \lim\limits_{\substack{t \to +\infty \\ t \in \hat{\mathcal{C}}_j}} x_j^t = \hat{x}$ for all $j \in \mathcal{P}$

Lemma: Assume the set $\mathcal{L}(x_0)$ is compact. Then there exists $\hat{x} \in \Re^n$ and $\hat{\mathcal{C}}_1 \subseteq \mathcal{C}_1$ such that

$$\lim_{\substack{t \to +\infty \\ t \in \hat{\mathcal{C}}_1}} \Delta_1^t = 0 \quad \text{and} \quad \lim_{\substack{t \to +\infty \\ t \in \hat{\mathcal{C}}_1}} x_1^t = \hat{x}.$$

Same argument as in Audet and Dennis, 1999.

$\Rightarrow$ We have an accumulation point for process 1.

Accumulation Point

Now we need to show that all the other processes also have $\hat{x}$ as an accumulation point.

Theorem 2: There exists $\hat{x}$ and, for each $j \in \mathcal{P}$, $\hat{\mathcal{C}}_j$ such that

$$\lim_{\substack{t \to +\infty \\ t \in \hat{\mathcal{C}}_j}} x_j^t = \hat{x} \text{ for all } j \in \mathcal{P}$$

Key: For each $\hat{t} \in \hat{\mathcal{C}}_1$ with $\hat{t} > t^*$, there is a corresponding time interval devoid of successful points on each of the other processes, so that they must both accept $x_1^{\hat{t}}$ as the best known point and have a number of contractions.

Search Directions

The pattern must be chosen so that it positively spans $\Re^n$. *See Lewis and Torczon, 1996.*

Defn: A set of vectors $\{d_1, \ldots, d_p\}$ *positively spans* $\Re^n$ if any vector $x \in \Re^n$ can be written as

$$x = \alpha_1 d_1 + \cdots + \alpha_p d_p, \quad \alpha_i \geq 0 \;\; \forall\, i.$$

That is, any vector can be written as a *nonnegative* linear combination of the basis vectors.

Fact: If $\{d_1, \ldots, d_p\}$ positively spans $\Re^n$, then for any $v \neq 0$, there exists d_i such that $d_i^T v < 0$.

FINAL RESULT

Theorem 3: Assume f is continuously differentiable. Then

$$\lim_{\substack{t \to +\infty \\ t \in \hat{\mathcal{C}}_i}} \nabla f(x_i^t) = 0.$$

Again, borrowing heavily from Audet and Dennis...

By MVT, for all $t \in \tau_i(\hat{\mathcal{C}}_i)$, $\exists\ \alpha_i^t \in [0,1]$ such that :

$$f(x_i^t) \le f(x_i^t + \Delta_i^t d_i) = f(x_i^t) + \Delta_i^t \nabla f(x_i^t + \alpha_i^t \Delta_i^t d_i)^T d_i,$$

$$\Rightarrow 0 \le \nabla f(x_i^t + \alpha_i^t \Delta_i^t d_i)^T d_i.$$

$$\Rightarrow 0 \le \nabla f(\hat{x})^T d_i \text{ for all } \quad i \in \mathcal{P}.$$

Since the d-vectors form a positive basis, that implies that $\nabla f(\hat{x}) = 0$.

Summary

Theorem 1: $\liminf_{t\to+\infty} \Delta_j^t = 0$ for all $j \in \mathcal{P}$.

Theorem 2: Assume $\mathcal{L}(x^0)$ is compact. Then there exists $\hat{x}$ and, for each $j \in \mathcal{P}$, $\hat{\mathcal{C}}_j$ such that

$$\lim_{\substack{t\to+\infty \\ t\in\hat{\mathcal{C}}_j}} x_j^t = \hat{x} \quad \text{for all } \; j \in \mathcal{P}$$

Theorem 3: Assume f is continuously differentiable. Then

$$\lim_{\substack{t\to+\infty \\ t\in\hat{\mathcal{C}}_j}} \nabla f(x_j^t) = 0 \quad \text{for all } \; j \in \mathcal{P}.$$

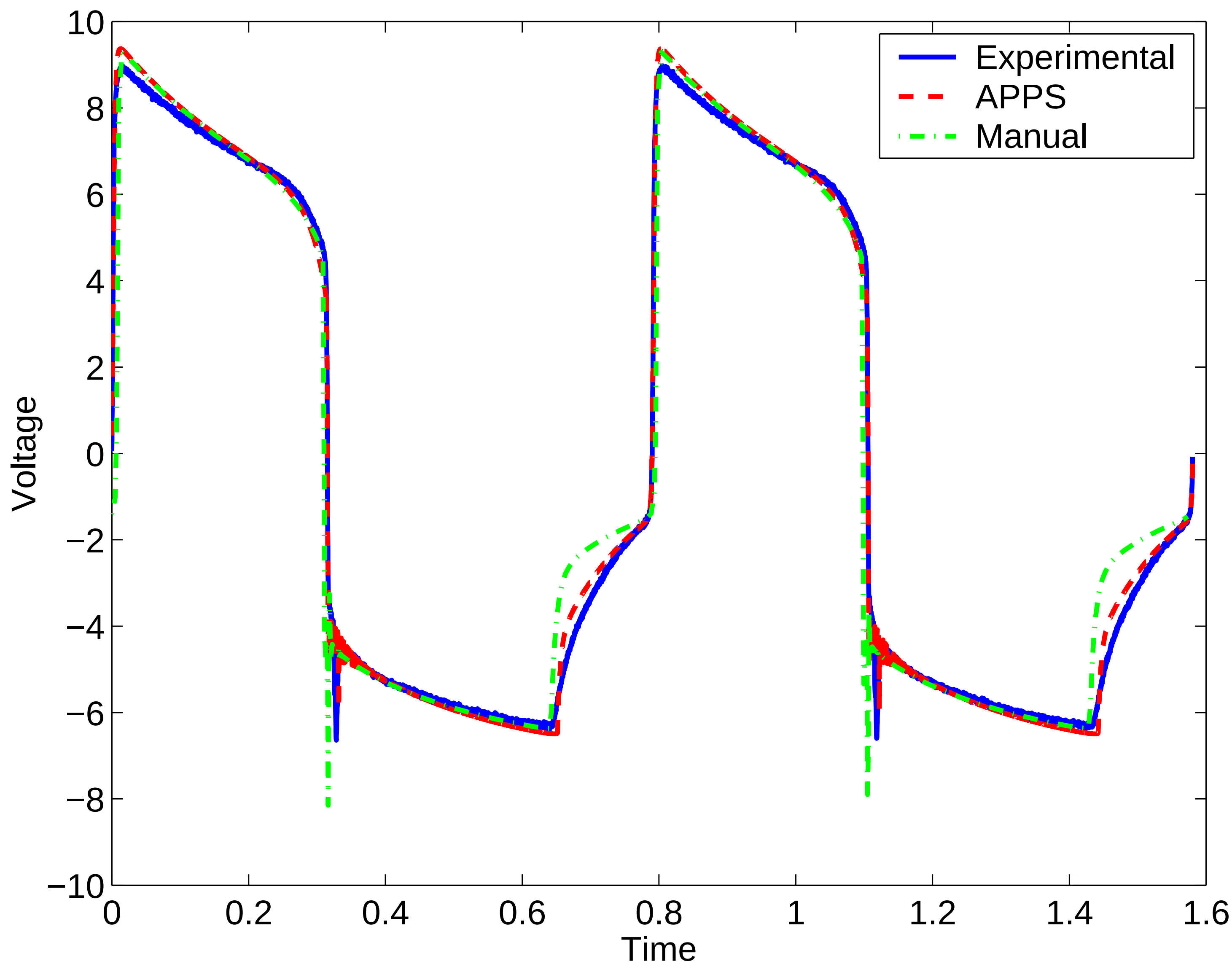
Electrical Circuit Simulation

- **Variables:** inductances, capacitances, diode saturation currents, transistor gains, leakage inductances, and transformer core parameters
- **Simulation Code:** SPICE3

$$f(x) = \sum_{t=1}^{N} \left(V_t^{\mathrm{SIM}}(x) - V_t^{\mathrm{EXP}} \right)^2 ,$$

$$\begin{aligned} x &= \text{17 unknown characteristics} \\ V_t^{\mathrm{SIM}}(x) &= \text{Simulation voltage at time } t \\ V_t^{\mathrm{EXP}} &= \text{Experimental voltage at time } t \\ N &= \text{Number of timesteps} \end{aligned}$$

Circuit Problem Results



WEB PAGE

http://csmr.ca.sandia.gov/projects/apps.html

APPSPACK

- PVM or MPI (Alton Patrick, NCSU)
- Bound constrained or unconstrained
- Function value cache (Patrick)
- Surrogate model (Sarah Brown, U. Wash.)

Research sponsored by the U.S. Department of Energy and the National Science Foundation.